Advanced Statistics

Analysis of Variance(ANOVA)

ANOVA Basics

- This technique is part of the domain called “Experimental Designs”.
- This helps in establishing in a precise fashion the Cause - Effect relation amongst variables.
- From the Statistical Inference Point of View, ANOVA is an extension of independent t test for testing the equality of two population means.
- When we have to compare more than two population means, we use ANOVA
- Typically, the null hypothesis(H_0) is as under:
- H_0 : $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \ldots\ldots = \mu_k$ for testing the equality of Population Means for k populations

ANOVA-One Way Classification

Assumptions involved in using ANOVA

- The samples drawn from different populations are independent and random.
- The response variables of all the populations are normally distributed.
- The variances of all the populations are equal.

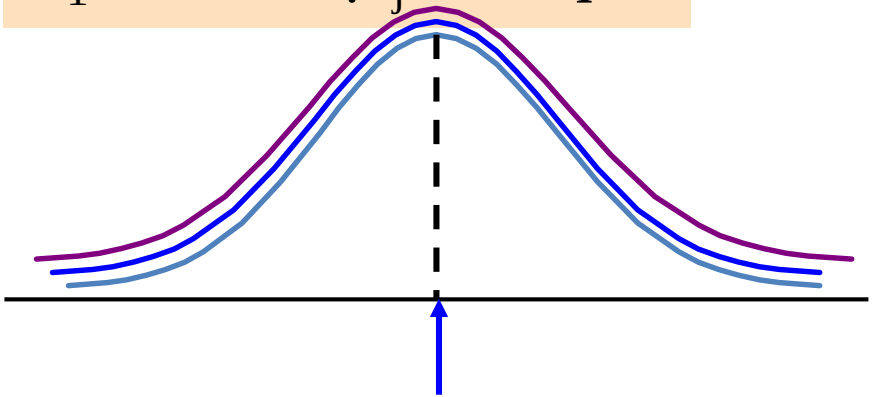
Hypotheses of One-Way ANOVA

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$
  - All population means are equal
- $H_1 :$ Not all of the population means are equal
  - For at least one pair, the population means are unequal

One-Way ANOVA
Null Hypothesis(H_0=True)

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

$$H_1 : \text{Not all } \mu_j \text{ are equal}$$

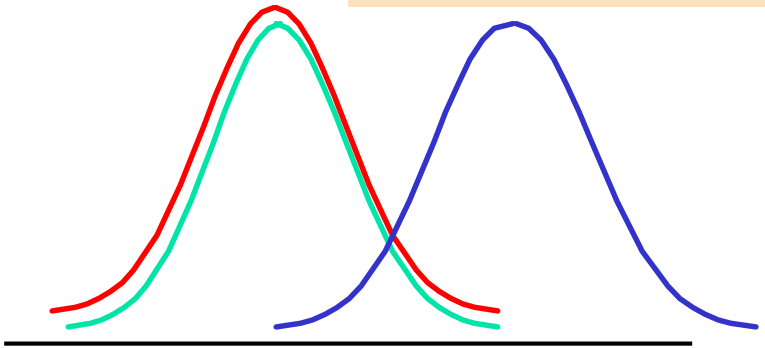


$$\mu_1 = \mu_2 = \mu_3$$

One-Way ANOVA
Alternative Hypothesis(H_1=True)

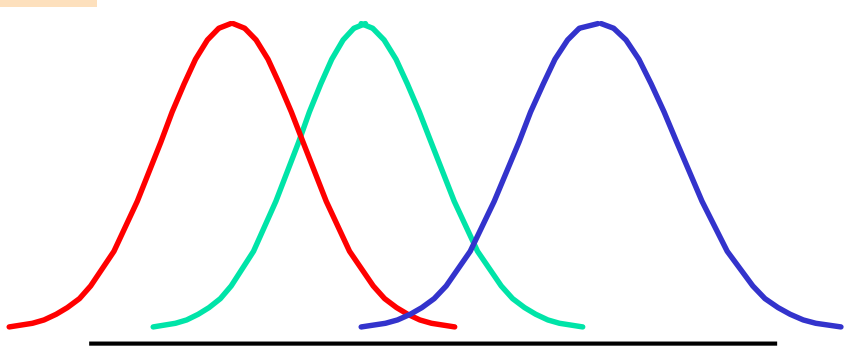
$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

$$H_1 : \text{Not all } \mu_j \text{ are equal}$$



$\mu_1 = \mu_2 \neq \mu_3$

or

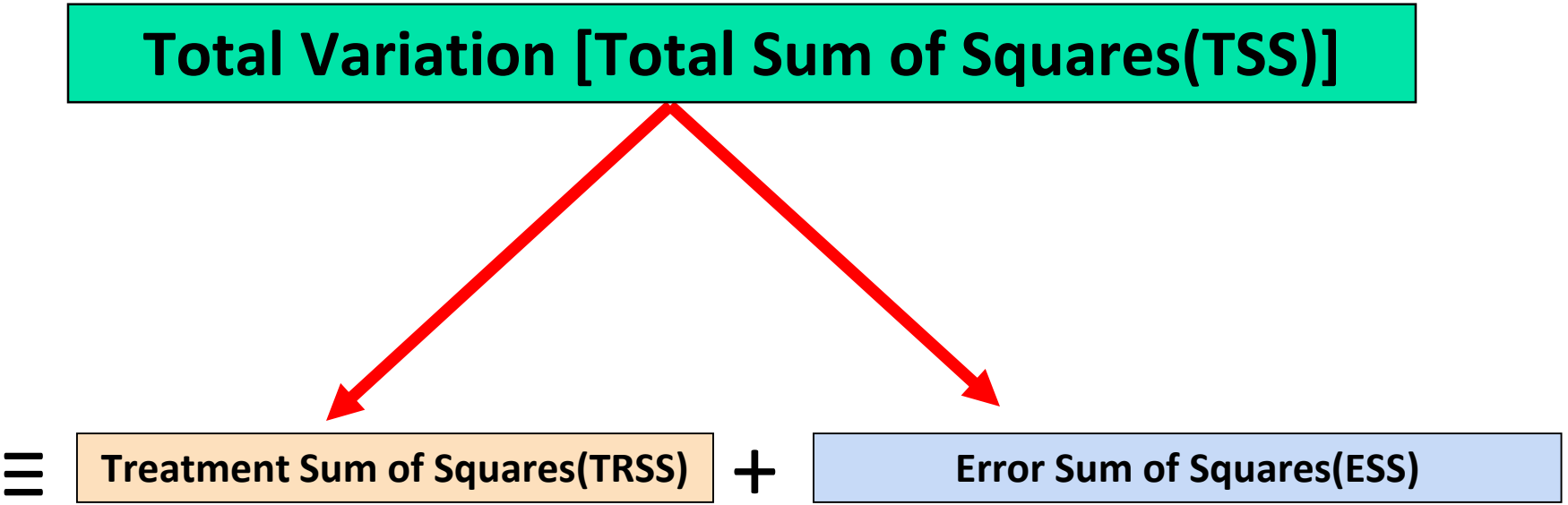


$\mu_1 \neq \mu_2 \neq \mu_3$

ANOVA Basics

- The beauty of ANOVA is that it performs the test of equality of more than two population means by actually analyzing the variance.
- In simple terms, ANOVA decomposes the total variation into components of variation. That is, explaining the changes in the response variable caused by these components.
- To put it succinctly, the total sum of squares is equal to the sum of squares due to causes.

Partition of Total Variation(Information Content



ANOVA-One Way Classification-Example

- A supermarket is interested in knowing whether it should go for a quarter-page, half-page, or a full-page advertisement for a Product.
- In order to choose the size of the advertisement that will bring in the most store traffic, the supermarket can use ANOVA technique.
- Here, you are trying to establish a cause-effect relationship between store traffic and the various sizes of advertisement.

ANOVA-One Way Classification

How One-Way Classification Works in Practice?

- **Total Sum of Squares ≡ Treatment Sum of Squares + Error Sum of Squares**.

- The word treatment is generic and as such may denote different methods, machines, different advertisement copy platforms, different strategies, different brands and the like.

- The variation in sum of squares of the response variable (dependent variable) is caused only by treatment and any thing unexplained by the treatment is attributed to error term.

Anova Output

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Design	3	2990.99	997.00	53.03	2.73E-13
Residuals	36	676.82	18.80		

Two-way ANOVA

- The two-way ANOVA compares the mean differences between groups that have been split on two independent variables (called factors).

- The primary purpose of a two-way ANOVA to understand if there is an interaction between the two independent variables on the dependent variable.

- For example, to understand if there is a relation between age and educational level on job experience amongst candidates, where age (25-30) and education level (undergraduate/postgraduate) are independent variables, and job experience is dependent variable.

Assumptions of two-way ANOVA

- Dependent variable should be measured at the continuous level.
- Two independent variables should each consist of two or more categorical, independent groups.
- There should be no significant outliers.
- Dependent variable should be approximately normally distributed for each combination of the groups of the two independent variables.

Difference between one-way ANOVA and two-way ANOVA

Basics for Comparison	One-way ANOVA	Two-way ANOVA
Meaning	One-way ANOVA is a hypothesis test, used to test the equality of three or more population means simultaneously using variance.	Two-way ANOVA is a statistical technique wherein, the interaction between factors, influencing variable can be studied
Independent Variable	One	Two
Compares	Three or more levels of one factor	Effect of multiple level of two factors
Number of Observation	Need not to be same in each group	Need to be equal in each group
Design of experiments	Need to satisfy only two principles	All three principles needs to be satisfied